Modification of antibiotic resistance genes and verification of protein function by antibiotic susceptibility testing

Iina Jormanainen

Master´s Thesis

University of Helsinki

November 2021

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Masters´s Programme
Faculty of Agriculture and Forestry	Master's Programme in Microbiology and Microbial Biotechnology

Tekijä — Författare — Author

Iina Jormanainen

Työn nimi — Arbetets titel — Title

Modification of antibiotic resistance genes and verification of protein function by antibiotic susceptibility testing

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	
Master's Thesis	November 2021	

Tiivistelmä — Referat — Abstract

The spread of antibiotic resistance in bacteria is a global problem. Horizontal gene transfer (HGT) is the main mechanism implicated in the spread of antibiotic resistance genes (ARG). This study is related to a doctoral thesis project that studies HGT in wastewater microbial community by conducting a microcosm experiment that uses Emulsion, Paired-Isolation and Concatenation PCR (epicPCR) to monitor the spread of ARGs between species.

The aim of this study was to introduce synthetic epicPCR primer binding sites inside various ARGs and to test the function of the encoded proteins. The goal was to maintain sufficient protein function, i.e., antibiotic resistance despite the modifications, which allows the further use of modified ARGs in microcosm experiment. The ARGs selected for modifications were *dfrB2*, *ermB*, *ermC*, *sul1* and *sul2*.

Sequence-based prediction method was applied to find regions that tolerate insertions inside the proteins encoded by ARGs. The modified ARGs carried in plasmid pUC19 were introduced to *Escherichia coli* DH5α, which was used as the host in antibiotic susceptibility testing. Antimicrobial gradient method was used to test the antibiotic susceptibility of the strains and to verify the function of the proteins.

Six ARGs modified in this study encoded for functional proteins that conferred antibiotic resistance while three modified ARGs did not. Two out of four proteins with insertions in predicted permissive stretches in the middle of a protein maintained their function. The six functional, antibiotic resistance conferring genes designed in this study can be used in further studies utilizing epicPCR. Based on the results of this study, sequence-based prediction method for finding permissive stretches seems useful, but it does not guarantee that the protein function is maintained.

Avainsanat — Nyckelord — Keywords

Antibiotic resistance, antibiotic susceptibility testing, protein engineering, sequence-based prediction

Säilytyspaikka — Förvaringsställe — Where deposited

https://ethesis.helsinki.fi/en/

Muita tietoja — Övriga uppgifter — Further information

Supervisors: Veera Partanen and Marko Virta

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Masters´s Programme
Maatalous-metsätieteellinen tiedekunta	Mikrobiologian ja mikrobibiotekniikan maisteriohjelma

Tekijä — Författare — Author

Iina Jormanainen

Työn nimi — Arbetets titel — Title

Antibioottiresistenssigeenien muokkaaminen ja proteiinien toiminnan testaus antibioottiherkkyysmäärityksellä

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	
Maisterintutkielma	Marraskuu 2021	

Tiivistelmä — Referat — Abstract

Antibioottiresistenssin leviäminen bakteereissa on maailmanlaajuinen ongelma. Horisontaalinen geeninsiirto on tärkein antibioottiresistenssigeenien leviämiseen liittyvä mekanismi. Tämä tutkimus liittyy väitöskirjatyöhön, jossa tutkitaan horisontaalista geeninsiirtoa jäteveden mikrobiyhteisössä mikrokosmoskokeen avulla. Kokeessa hyödynnetään epicPCR-menetelmää (Emulsion, Paired-Isolation and Concatenation PCR) antibioottiresistenssigeenien leviämisen seurannassa lajien välillä.

Tämän tutkimuksen tarkoituksena oli lisätä synteettiset epicPCR-alukkeiden sitoutumiskohdat usean eri antibioottiresistenssigeenin sisään ja testata geenien koodaamien proteiinien toiminta. Tavoitteena oli ylläpitää riittävä proteiinin toiminta eli antibioottiresistenssi muokkauksista huolimatta, mikä mahdollistaa muokattujen antibioottiresistenssigeenien jatkokäytön mikrokosmoskokeessa. Muokattaviksi valitut geenit tässä työssä olivat *dfrB2*, *ermB*, ermC, *sul1* ja *sul2*.

Sekvenssipohjaista ennustamismenetelmää sovellettiin insertion sallivien paikkojen etsimisessä antibioottiresistenssigeenien koodaamien proteiinien sisältä. Muokatut geenit siirrettiin vektorissa pUC19 *Escherichia coli* DH5α -kantaan, jota käytettiin isäntänä antibioottiherkkyystestauksessa. Kantojen antibioottiherkkyyden testaamiseen ja proteiinien toiminnan tarkistamiseen käytettiin gradienttiliuskamenetelmää.

Kuusi tässä tutkimuksessa muokatuista antibioottiresistenssigeeneistä tuotti toiminnallisen, antibioottiresistenssin antavan proteiinin, kun taas kolme muokatuista geeneistä eivät. Kaksi neljästä proteiinista, joille oli tehty insertio sen sallivaksi ennustettuun paikkaan, säilytti toiminnallisuutensa. Tässä tutkimuksessa suunniteltua kuutta toiminnallista, antibioottiresistenssin tuottavaa geeniä voidaan käyttää epicPCR:ää hyödyntävissä jatkotutkimuksissa. Tämän tutkimuksen tulosten perusteella sekvenssipohjainen ennustamismenetelmä sallivien paikkojen löytämiseksi vaikuttaa käyttökelpoiselta, mutta se ei kuitenkaan takaa proteiinin toiminnallisuuden säilymistä.

Avainsanat — Nyckelord — Keywords

Antibioottiresistenssi, antibioottiherkkyysmääritys, proteiinitekniikka, sekvenssipohjainen ennustaminen

Säilytyspaikka — Förvaringsställe — Where deposited

https://ethesis.helsinki.fi/

Muita tietoja — Övriga uppgifter — Further information

Ohjaajat: Veera Partanen ja Marko Virta

INTRODUCTION

Antibiotic resistance of bacteria is one of the biggest problems the world is facing today (World Health Organization, 2020). Antibiotic resistance has existed naturally before human antibiotic use (Hughes & Datta, 1983) and even now it is present in natural environments (Allen et al., 2010). Many human, animal and natural environments are involved in the emergence, acquisition and spread of antibiotic resistance (Hernando-Amado et al., 2019). Irresponsible use of antibiotics in agriculture and human healthcare has caused the spreading of antibiotic resistance to increase. Other factors contributing to the spread are for example the co-selection of antibiotic resistance genes with heavy metals and biocides and the effects of global warming such as the increase of space where bacteria, humans, animals, and vector species can interact. The increasing spread of antibiotic resistance endangers both human and animal health as well as food security (World Health Organization, 2020). The spread of antibiotic resistance in human pathogens has led to antibiotics losing their effectiveness, making injuries and common infections such as pneumonia, tuberculosis, salmonellosis, and gonorrhea harder to treat. As a result, medical costs become higher, and mortality increases.

Clinical and agricultural use of antibiotics leads to the presence of antibiotics in soil and aquatic environments (Allen et al., 2010). One reason for this is that the antibiotics consumed by humans or animals are largely excreted from the body (Thiele-Bruhn, 2003). The persistence of antibiotics in the environment creates a pressure that selects for antibiotic resistance in bacteria (Allen et al., 2010; Hiltunen et al., 2017). Big part of the problem is that many of the known antibiotic resistance genes (ARGs) are found in mobile genetic elements such as plasmids and transposons and therefore ARGs can effectively spread between bacterial strains and species. A main mechanism for the spreading of antibiotic resistance is horizontal gene transfer, the exchange of genetic material between bacterial cells that allows bacteria to acquire new genes from the environment.

Wastewater treatment plants (WWTPs) are possible hotspots for the spread of ARGs and bacteria (Rizzo et al., 2013). Because bacteria from different environmental sources enter the WWTPs and are there in close contact, it is possible that ARGs are transferred from environmental bacteria to a pathogen, or vice versa. Previous studies suggest that ARGs are transferred between species in

WWTPs, though the treatment process mainly succeeds in decreasing the host range of ARGs (Hultman et al., 2018). In WWTPs bacteria and antibiotics at sub-inhibitory concentrations are simultaneously present, which creates a suitable environment for HGT and thus for the spreading of antibiotic resistance (Rizzo et al., 2013). However, the factors and mechanisms that are responsible for maintaining and selecting antibiotic resistance in wastewater are not yet well understood.

This Master's thesis is related to a Doctoral thesis project of Veera Partanen, which aims to monitor the transfer of ARGs in wastewater microbial communities in experimental conditions. To follow the spread of ARGs in microbial communities a method called Emulsion, Paired-Isolation and Concatenation Polymerase Chain Reaction (epicPCR) (Spencer et al., 2016) will be applied. The method reveals the current hosts of the ARG by linking it with 16S rRNA gene on single cell level. For Doctoral thesis project, epicPCR primer binding sites need to be added to several different ARGs and these primer binding sites need to be the same. Having the same epicPCR primer binding sites in different ARGs allows the amplification of several ARGs in one reaction. If the primer binding sites were specific to each ARG, every gene would have to be individually analyzed, making the epicPCR more laborious. ARGs with primer binding sites for epicPCR will also be tagged with a unique barcode sequence, which tells the donor of the ARG, while epicPCR linking reveals the current host.

Adding amino acids in the middle of a protein while retaining protein function requires the identification of a permissive site. A permissive site is a place in a protein which tolerates relatively large insertions without a loss of protein function (Oesterle et al., 2017). One approach for permissive site identification is to use a pentapeptide scanning mutagenesis, which is a technique that utilizes transposons (for example in Goodale et al. 2020). Another approach for finding permissive sites is to use sequence-based prediction (Oesterle et al., 2017). This relatively new method relies on the hypothesis that protein's amino acid sequence information is enough to reveal stretches that tolerate insertions (Burg et al., 2016; Oesterle et al., 2017). Regions that have low conservation and are variable in length between homologs are thought to be less likely relevant for the protein function. Length variable regions contain insertions or deletion (indels) and indels often occur in a loop structure on the surface of a protein (Chang & Benner, 2004). Sequence-based prediction of permissive sites was chosen for this study because the novel method claimed to be less laborious than the pentapeptide scanning mutagenesis. Also, the pentapeptide scanning

mutagenesis leaves a transposase recognition site flanking the inserted sequence, which may lead to the loss of protein function (Billerbeck et al., 2013).

EpicPCR is a relatively new method and is not known to have been used to amplify long sequence fragments. Successful amplification with epicPCR has been reported when a sequence to be linked to 16s rRNA were less than 300 nucleotides in length (Cairns et al., 2018; Hultman et al. 2018; unpublished data). Therefore, it is necessary to determine whether the binding site of the epicPCR primer can be placed in the middle of the gene of interest, making the length of the sequence to be amplified and linked more advantageous for epicPCR. In addition, it is necessary to find out whether the primer binding sites can be placed inside the start and stop codons of a gene. This would allow primer binding sites to be selected with the ARG even if the gene changes its genetic environment.

The aim of this study was to introduce synthetic primer binding sites and a barcode sequence inside various ARGs and to test the function of the encoded proteins. The function of the protein, i.e., the antibiotic resistance, was to be maintained despite the modifications. The required features for the ARGs to be designed were that they needed be amplifiable by epicPCR and selectable by the corresponding antibiotic in experimental conditions. To ensure successful amplification by epicPCR, addition of the epicPCR primer binding site sequence to the middle of a gene in a region predicted as permissive was explored for some the genes selected for this study. However, epicPCR itself was not performed in this study and thus the epicPCR amplifiability of the genes was not tested. The function of the proteins was verified with antibiotic susceptibility testing where the transformant strain carrying the modified genes were compared to strains carrying wild type genes.

MATERIALS AND METHODS

Genes and proteins

The ARGs for this study were searched from a set of ARGs located on plasmids available for the Molecular Environmental Biosciences research group, in which this study was conducted (Table 1). The requirement for these thirteen genes was that the gene needed to be beneficial to the host only and not to other microbes in the environment. Of these thirteen ARGs, six genes, *dfrB2*, *ermB*, *ermC*, *ermF*, *sul1* and *sul2*, were selected for this study (Table 2). The genes were selected because a 3D

model of the structure of the protein encoded by the gene is published in a public database. 3D models were searched against UniProtKB/SwissProt database (Boutet et al. 2007) and Protein Data Bank (Berman et al., 2000) based on amino acid sequence using NCBI BLASTp (Altschul et al. 1990) with default settings.

Table 1. **A list of antibiotic resistance genes located on plasmids.** The GenBank accession numbers refer to the plasmids on which the gene is located, except for *ermF* gene where the accession number refers to the gene only. Genes of the same name are marked with different numbers in the superscript.

Gene	Plasmid	Accession number
dfrA	pKJK5	AM261282.1
dfrB2	R388	NC_028464.1
ermB	pAMbeta1	NC_013514.1
ermC	pE194	NC_005908.1
ermF	pVA831 (pBF4)	M14730.1
floR	pAB5S9	NC_009476.1
sul1	pKJK5	AM261282.1
sul2^{1}	pAB5S9	NC_009476.1
sul2^{2}	RSF1010	M28829.1
tetA^{1}	pKJK5	AM261282.1
tetA^{2}	RP4	X75761.1
tetC	pRAS3.3	NZ_KJ909291.1
tetY	pAB5S9	NC_009476.1

Table 2. **Data of antibiotic resistance genes used in this study.** An antibiotic against which the gene confers resistance, size of the gene in base pairs (bp) and proteins encoded by antibiotic resistance genes as well as their UniProtKB accession numbers.

Gene	Resistance against	Size (bp)	Protein	Accession number
dfrB2	trimethoprim	236	Dihydrofolate reductase	P00384
ermB	macrolide-lincosamide-streptogramin B resistance (erythromycin)	737	23S rRNA (adenine(2058)-N(6))-methyltransferase	P0A4D5
ermC	macrolide-lincosamide-streptogramin B resistance (erythromycin)	734	23S rRNA (adenine(2058)-N(6))-methyltransferase	P02979
ermF	macrolide-lincosamide-streptogramin B resistance (erythromycin)	798	23S rRNA adenine(2058)-N(6))-methyltransferase	P10337
sul1	sulfonamide	839	Dihydropteroate synthase type 1	P0C002
sul2[1]	sulfonamide	815	Dihydropteroate synthase type 2	P0AC11

Permissive stretch search

In this work, a previously described method for finding a permissive stretch based on amino acid sequence was applied (Oesterle et al. 2017). Proteins of interest were searched for permissive stretches by aligning the amino acid sequence of each protein with the amino acid sequence of 4-6 homologous proteins. The homologs were retrieved from Universal Protein Resource Knowledgebase (Bateman et al., 2021) using Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) (Boratyn et al. 2012). Criteria for homologs selected for multiple sequence alignment (MSA) were similarity of 30% to 70% in amino acid sequence and a query coverage greater than 80%.

The multiple sequence alignment was performed using the online version of Clustal Omega program with default parameters (November 17th–18th, 2020) (Sievers et al., 2011). From MSA data, gaps were searched to identify a region permissive to insertion. A possible permissive stretch was identified as a gap in the MSA along with residues flanking the gap. In this work, the stretches were searched for within the region of the first 100 amino acids so that the sequence to be amplified in epicPCR would be less than 300 base pairs. The location of the permissive stretch was further

evaluated by examining the 3D structural data of the protein. The aim was to find stretches located in a flexible loop region.

For protein encoded by *ermB,* DELTA-BLAST yielded homologs, but the homologs did not reveal gaps in the MSA. Therefore, for this protein the homologs were retrieved using protein-protein BLAST with default parameters as they revealed gaps in the MSA. For protein encoded by *dfrB2* permissive stretch search was not done because inserting the epicPCR primer binding site in the middle of the gene was not necessary. *DfrB2* is relatively small (236 bp) and hence the epicPCR product would be within the desired length (less than 300 bp) when the whole gene is amplified.

Modified genes

Three types of genes were designed in this study (Fig. 1.). Type A gene included an ARG sequence, forward primer binding site for epicPCR after the start codon, a barcode sequence following the forward primer binding site for epicPCR and a reverse forward primer binding site for epicPCR at the end of the gene before stop codon. In type B gene, the reverse forward primer binding site for epicPCR was placed in the middle of the protein in a region identified as permissive. For type C gene no modifications were made. Type C is therefore also referred as a wild type gene. For each gene, Shine-Dalgarno sequence (AGGAGG) was added upstream of the start codon to ensure ribosome binding, which allows gene expression in the host plasmid. Five bases (CAGCT) were added between Shine-Dalgarno sequence and start codon. These five bases were chosen since the same set precedes the start codon of *lacZ* gene before Shine-Dalgarno sequence in the used vector plasmid pUC19.

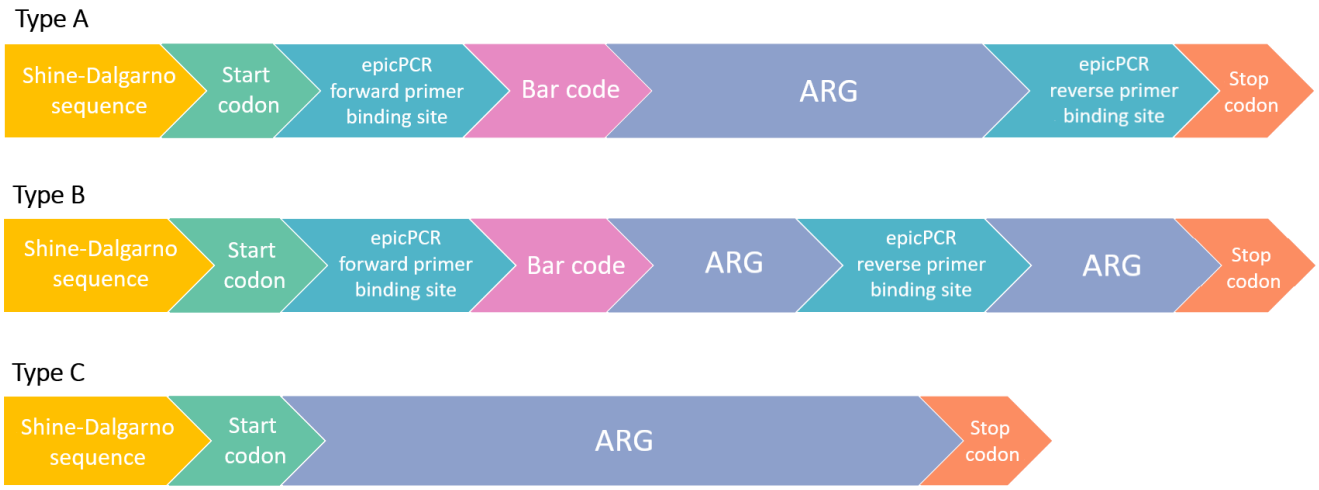

Figure 1. **Illustration of designed inserts.** Types of modifications: **(A)** epicPCR forward primer binding site and barcode added after the start codon, epicPCR reverse primer binding site added to the end of the gene before stop codon, **(B)** epicPCR forward primer binding site and barcode added after the start codon, epicPCR reverse primer binding site added to the middle of the gene **(C)** wild type gene; no epicPCR primer binding sites added.

The adding of epicPCR primer binding sites and barcode (Table 3) to the ARG sequence required addition of extra bases to maintain correct reading frame for the gene. The bases were selected so that the amino acids encoded by the codons would be as small and harmless to the protein structure as possible. Amino acids known to have a key role in protein structure such as cysteine (Matsumura et al., 1989) and proline (MacArthur & Thornton, 1991) were avoided, as well as other amino acids with long side chains or side chains containing sulfur. As the insertions in predicted permissive stretches for type B genes were to be done on surface-exposed loops, the amino acids encoded by epicPCR primer binding sites were confirmed to be mainly hydrophilic amino acids, which are prone to exist on protein surfaces.

Table 3. **Nucleotide and amino acid sequences of insertions to antibiotic resistance genes.** Forward epicPCR primer binding site is in blue, reverse epicPCR primer binding site in green, barcode sequence in pink and additional bases in yellow.

	Nucleotide sequence	**Amino acid sequence**
EpicPCR forward primer binding site and barcode	GGTCGTGAGCACCTAGGGTCTCATGCCATT	GREHLGSHAI
EpicPCR reverse primer binding site	GGGCAGAGCCTCAGAACACTT	GQSLRTL

For the proteins encoded by genes *ermB*, *ermC*, *sul1* and *sul2* all three gene types were explored. *ErmF* was excluded from further experiments because of its difficulties to be expressed in *Escherichia coli* (Rasmussen et al. 1986), which was the chosen host for both cloning and expression. In type B, epicPCR reverse primer binding site was inserted to a site identified as permissive. The insertion was done either by inserting the new sequence without deleting residues or by deleting a residue or more in the process. Amino acid sequences highlighting the permissive stretches and insertions are shown in Table 4. The naming of the genes in this study is as following: the letter a, b or c following the original gene name indicates the type of modification done to the gene (Fig. 3). The name of an amino acid residue indicates the insertion site in the protein encoded by the gene. For example, *ermB_b_91N* is *ermB* gene with type B modification where epicPCR reverse primer binding site is added immediately after residue 91N.

The designed inserts were placed in a plasmid pUC19 (Yanisch-Perron et al., 1985) at HindIII/BamHI cloning site. The plasmids containing the designed inserts were ordered from Genscript Biotech, Netherlands and they were delivered in lyophilized form. The lyophilized plasmids were dissolved into 20 µl of sterile water according to the manufacturer's instructions to prepare them for transformation. The plasmid DNA concentrations were determined using Qubit 4 Fluorometer (Invitrogen by Thermo Fisher Scientific) (data not shown) and appropriate solutions were made for transformation.

Table 4. **Overview of insertions in predicted permissive stretches.** The added sequence to the target protein is the amino acid sequence corresponding to the epicPCR reverse primer binding site (Table 3). If a base/residue was deleted from the original sequence during insertion, it is underlined. The added bases/residues are in boldface.

Protein	**Insertion site**	**Original nucleotide sequence of the permissive stretch**	**Nucleotide sequence after insertion**	**Original amino acid sequence of the permissive stretch**	**Amino acid sequence after insertion**
23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermB*)	91N	AAC<u>AAA</u>CAG	AAC**GGGCAGAGCCTCAGAACACTT**CAG	N<u>K</u>Q	N**GQSLRTL**Q
23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermC*)	72L	CTT<u>GTT</u>GAT	CTT**GGGCAGAGCCTCAGAACACTT**GAT	L<u>V</u>D	L**GQSLRTL**D
Dihydropteroate synthase type 1 (*sul1*)	72L	CTGTCC	CTG**GGGCAGAGCCTCAGAACACTT**TCC	LS	L**GQSLRTL**S
Dihydropteroate synthase type 2 (*sul2*)	75L	CTC<u>AAG</u>GCA	CTC**GGGCAGAGCCTCAGAACACTT**GCA	L<u>K</u>A	L**GQSLRTL**A

Bacterial strains and growth conditions

Escherichia coli DH5α strain was used as a cloning and expression host. The bacteria were grown on Luria-Bertani (LB) agar or LB broth medium at +37°C, the latter with shaking (220 rpm). For transformants with pUC19 plasmid 100 µg/ml of ampicillin was added for selection. Ampicillin was the selective agent on the plates as pUC19 plasmid contains an ampicillin resistance gene.

Preparation of competent cells

One colony from an overnight pure culture plate of *E. coli* DH5α was inoculated into a 50 ml LB broth flask. The flask was incubated overnight at +37 °C with shaking and on the following day 25 ml of the culture was inoculated into 500 ml of LB broth. The cells were grown with vigorous shaking at +37 °C to an OD_{580} of 0.4. Immediately after this the culture flask was chilled on ice approximately for 25 min. The cooled cell culture was transferred to sterile centrifuge bottle and 275 ml of LB broth was added to reach a final volume of 800 ml. The cells were centrifuged at 1000 x g for 15 min at 4 °C. The supernatant was removed, and the cells were resuspended in 800 ml of ice-cold sterile water and centrifuged at 1000 x g for 20 min at 4 °C. The pelleted cells were resuspended in 250 ml of ice-cold 10 % (v/v) glycerol and centrifuged as above. Again, the supernatant was removed, and the pelleted cells were resuspended in 10 ml of 10 % (v/v) glycerol and centrifuged as above. The final suspension of cells was done in 800 µl of ice-cold 10 % (v/v) glycerol. The OD_{600} value of the 1:100 suspension was 0.375. The suspension was divided in 40 µl aliquots and the tubes were stored at -80 °C until use.

Transformation

The transformation was carried out by electroporation. The electroporation was performed according to the New England Biolabs (2019) protocol with minor modifications. Plasmid pUC19 was used as a positive and PCR grade water as negative electroporation control. 1 µl of plasmid DNA solution (10 pg/ µl) or PCR grade water was mixed with 40 µl of electrocompetent cells. The cell solution was transferred to a pre-chilled electroporation cuvette (Sigma-Aldrich) with 0.1 cm gap. The electroporation was carried out with Bio-Rad Gene Pulser. Electroporation conditions were 200 Ω, 25 µF and 2.00 kV.

Screening of transformants

After the electroporation, the cells were cultured on selective media. 100-fold and 1000-fold dilutions were made of the cell suspension and 100 µl of both dilutions were plated. The undiluted cell suspension was centrifuged (6000 x g, 5 min) and resuspended in residual SOC (Super Optimal broth with Catabolite repression) solution and plated to ensure the yield of transformants. In addition, 100 µl of 1000-fold dilution was plated on LB agar plate without ampicillin to control whether the cells survived the electroporation. 100-fold dilution was made of the negative control after electroporation and 100 µl was plated on LB agar plates with and without ampicillin to test its selectiveness.

The plates were screened for transformants, and three colonies were selected for further testing. The colonies were checked by colony PCR for correct insert size. The used primer pair (fwd 5' GTGAGTTAGCTCACTCATTAGGC 3', rev 5' CCAACTTAATCGCCTTGC 3', Metabion international AG) was specific to plasmid pUC19. The predicted PCR product lengths are presented in Table 5.

Table 5. **Colony PCR product sizes.**

Plasmid	**Predicted product size (bp)**
pUC19	246
pUC19-*dfrB2_a*	527
pUC19-*dfrB2_c*	476
pUC19-*ermB_a*	1028
pUC19-*ermB_b_91N*	1025
pUC19-*ermB_c*	977
pUC19-*ermC_a*	1025
pUC19-e*rmC_b_72L*	1022
pUC19-*ermC_c*	974
pUC19-*sul1_a*	1130
pUC19-*sul1_b_72L*	1130
pUC19-*sul1_c*	1067
pUC19-*sul2_a*	1106
pUC19-*sul2_b_75L*	1103
pUC19-*sul2_c*	1055

For strains carrying plasmids pUC19, pUC19-*dfrB2_a* and pUC19-*dfrB2_c* PCR was performed in a volume of 20 µl which contained 0.4 U of Phusion™ High-Fidelity DNA Polymerase (Thermo Fischer Scientific), dNTP Mix (0.2 mM of each) 4 µl of 5-fold Phusion™ HF Buffer (Thermo Fischer Scientific), 0.5 µM of each of the two primers and template DNA from transformant colonies growing on LB

agar plates with ampicillin. The colony PCR was performed using the following program: initial denaturation at 98 °C for 2 min, followed by 30 cycles of 98 °C for 10 s, 59,5 °C for 30 s, and 72 °C for 20 s and final extension step at 72 °C for 5 min.

For the rest of the strains, colony PCR with Phusion™ High-Fidelity DNA Polymerase did not yield products. Therefore the colony PCR for the strains with plasmids pUC19-*ermB_a*, pUC19-*ermB_b_91N*, pUC19-*ermB_c*, pUC19-*ermC_a*, pUC19-*ermC_b_72L*, pUC19-*ermC_c*, pUC19-*sul1_a*, pUC19-*sul1_b_72L*, pUC19-*sul1_c*, pUC19-*sul2_a*, pUC19-*sul2_b_75L* and pUC19-*sul2_c* was performed in a volume of 25 µl which contained 1.25 U of DreamTaq DNA Polymerase (Thermo Fischer Scientific), dNTP Mix (0.2 mM of each) 2,5 µl of 10-fold DreamTaq Buffer (Thermo Fischer Scientific), 0.5 µM of each of the two primers and template DNA from transformant colonies growing on LB agar plates with ampicillin. The colony PCR was performed using the following program: initial denaturation at 95 °C for 3 min, followed by 30 cycles of 95 °C for 30 s, 59,5 °C for 30 s, and 72 °C for 60 s and final extension step at 72 °C for 5 min.

The colony PCR was performed in the Bio-Rad C1000 Touch™ Thermal Cycler instrument. The PCR products were analyzed by gel electrophoresis (E-gel iBase™, Invitrogen by Thermo Fischer Scientific) on a 2 % agarose E-gel (G401002, Invitrogen by Thermo Fischer Scientific) with GeneRuler 1 kb DNA Ladder (SM0311, Thermo Fischer Scientific) or GeneRuler 50 bp DNA Ladder (SM0371, Thermo Fischer Scientific) for evaluation of the fragment sizes.

Preparation of glycerol stocks

The PCR-verified transformants were inoculated in 5 ml LB broth with 100 µg/ml ampicillin and glycerol stocks of each transformant were made after an overnight incubation at +37 °C. The stocks were prepared by mixing 0.5 ml of 85 % (v/v) sterile glycerol with 1 ml of bacterial cell culture in an Eppendorf tube. The tubes were stored at -80 °C.

Plasmid DNA extraction and analysis of sequencing data

In the agarose gel analysis, the PCR product sizes were observed to be approximately as expected. However, the products seemed slightly bigger than what was calculated. For this reason, one plasmid containing *ermB_b_91N* gene was sequenced to confirm that the transformants carry the

right plasmids. Plasmid DNA extraction was done for *E. coli* DH5α transformant carrying pUC19 plasmid with *ermB_b_91N* gene. The extraction was done using Monarch® Plasmid Miniprep Kit (T1010S, New England BioLabs) according to manufacturer's instructions (International.neb.com, 2015). The DNA was eluted to 30 µl of Monarch DNA Elution Buffer. The plasmid DNA concentrations were determined using Qubit 4 Fluorometer (Invitrogen by Thermo Fisher Scientific).

The extracted plasmid pUC19-*ermB_b_91N* was sent for two-way Sanger sequencing along with original pUC19-*ermB_b_91N* provided by Genscript, USA. The sequencing was performed, and sequence data was provided by FIMM Sequencing. The same primer pair that was used in colony PCR was used in Sanger sequencing as well. The sequencing data was analyzed with NCBI BLASTn (Altschul et al., 1990).

Antibiotic susceptibility testing

Antimicrobial gradient method was used to test the antibiotic susceptibility in form of minimal inhibitory concentration (MIC) values for transformant strains. The testing was performed on Mueller-Hinton (MH) agar with ampicillin (100 µg/ml) using Liofilchem® MIC test strips (Liofilchem, Italy) according to the manufacturer's instructions. Antibiotics used in the testing were trimethoprim, erythromycin, and sulfamethoxazole. Transformant strains from glycerol stocks were precultured on LB agar with ampicillin (100 µg/ml) to get single colonies. Colonies were suspended in 2 ml of sterile 0.85 % (w/v) NaCl solution corresponding to 0.5 McFarland standard. The bacterial suspension was plated on MH agar with ampicillin (100 µg/ml) with a cotton swab to achieve a confluent lawn of bacteria. The MIC test strip was applied on agar surface and the plate was incubated for 16-20 hours in + 37 °C. After incubation, MIC value was read where the visible inhibition zone intersected the strip. An exception for the manufacturer's instructions was the addition of ampicillin to MH agar media, which was done to avoid the loss of plasmid from the transformed strains. A summary of the antibiotic susceptibility testing protocol is outlined in Fig 2.

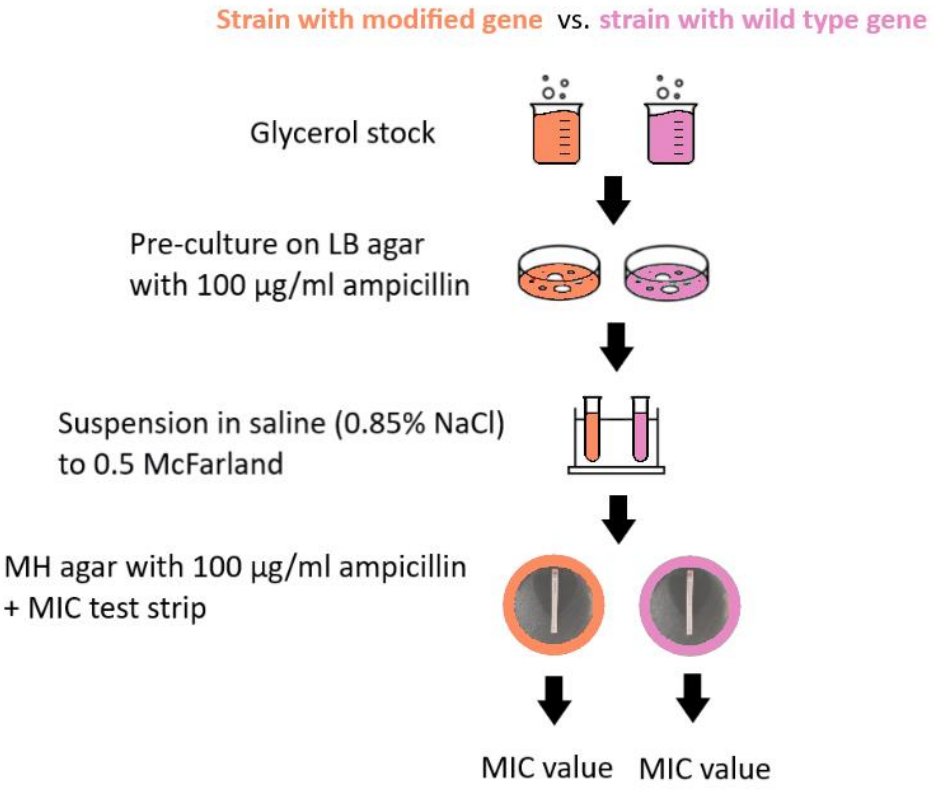

Figure 2. **A summary of the antibiotic susceptibility testing protocol.** Bacterial strain carrying modified ARG is illustrated in orange and wild type strain in pink. Abbreviations: LB = Luria-Bertani, MH = Mueller-Hinton, MIC = minimal inhibitory concentration.

The antibiotic susceptibility testing of all strains was done twice and the MIC values of the two test rounds were compared. A third testing was performed if the MIC values of the two tests differed considerably from each other. Separately conducted antibiotic susceptibility testing experiments aimed to be independent from each other. For each round of testing, media was prepared separately. Each test round started with collecting the transformant cells were from a glycerol stock for the pre-culture. The same glycerol stock tube of each transformant strain was used in every test. *E. coli* DH5α (pUC19) was used as a negative control for the three tested antibiotics. *E. coli* DH5α was used to control the ampicillin selection on the plates.

RESULTS

Identification of permissive stretches

Gaps in the MSA were observed for all the studied proteins encoded by genes *ermB, ermC, ermF, sul1* and *sul2*. One or more permissive stretch within the region of the first 100 amino acids was identified for all the proteins (Table 6). One stretch was chosen for each protein to be the site of insertion. The stretches chosen for insertion were located in flexible loop regions. However, in the absence of more suitable alternative, for proteins encoded by *sul1* and *sul2,* the stretch was located at the end of an alpha helix just before a loop. If more than one suitable stretch was identified for a protein, the stretch that was relatively least conserved was chosen to be the insertion site. The locations of the chosen permissive stretches for each protein are shown in figures of 3D protein structures (Fig 3.)

Table 6. **Results of permissive stretch search.** The location of permissive stretch is marked as the first flanking amino acid of the stretch. In the column reporting lengths of permissive stretches in amino acids, the amino acid in parentheses indicates the location of the stretch.

Protein	**No of permissive stretches identified within the region of the first 100 amino acids**	**Length of permissive stretch in amino acids**	**Location of the chosen permissive stretch**
23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermB*)	2	3 (73L) 3 (91N)	91N
23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermC*)	2	3 (27R) 3 (72L)	72L
23S rRNA adenine(2058)-N(6))-methyltransferase (*ermF*)	3	3 (33N) 3 (78A) 5 (97P)	-
Dihydropteroate synthase type 1 (*sul1*)	1	3 (72L)	72L
Dihydropteroate synthase type 2 (*sul2*)	3	3 (26A) 4 (75L) 3 (97S)	75L

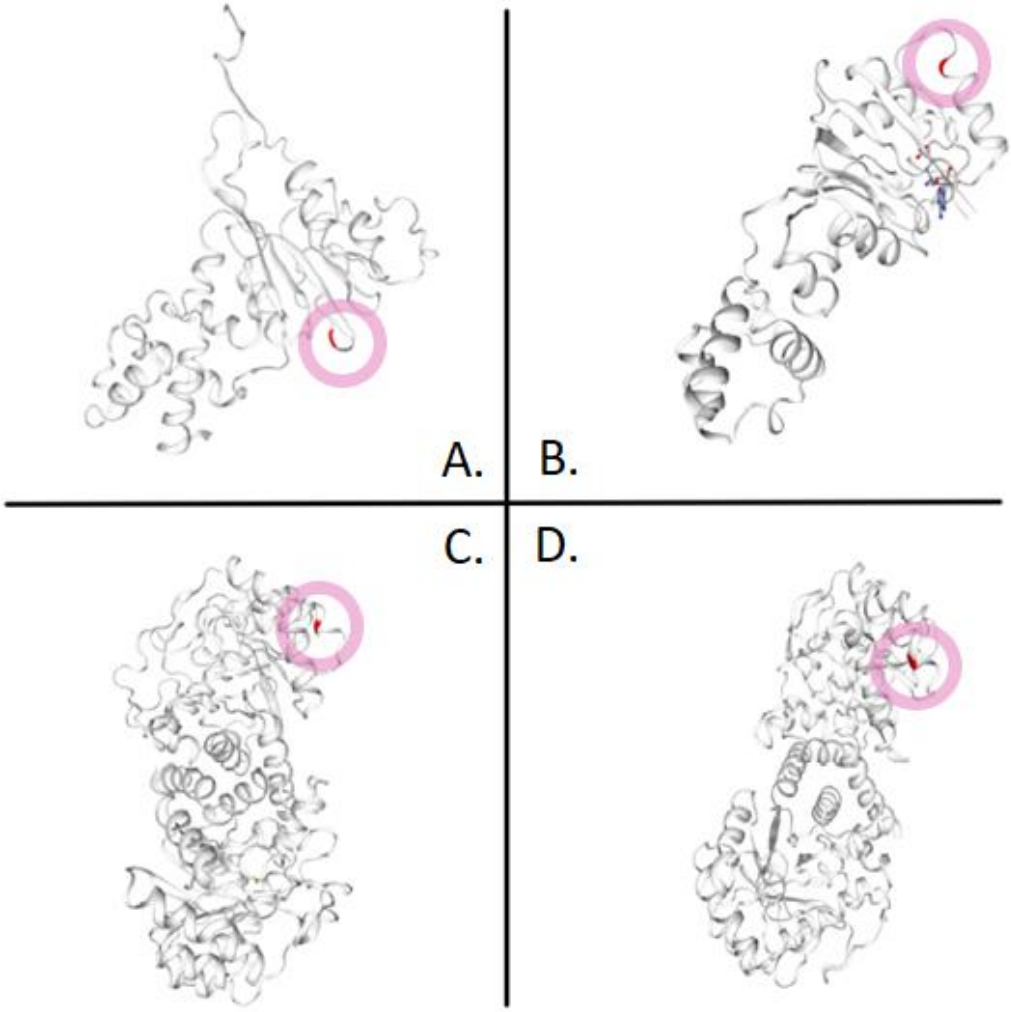

Figure 3. **3D structure models displaying the locations of the chosen permissive stretches.** The first flanking residue of the stretch is highlighted in red. **(A)** 23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermB*) (Bhujbalrao & Anand, 2019). **(B)** 23S rRNA (adenine(2058)-N(6))-methyltransferase (*ermC*). The 3D structure model contains a sinefungin ligand, visible in blue and red (Schluckebier et al., 1999). **(C)** Dihydropteroate synthase type 1 (*sul1*) (Yun et al., 2012). **(D)** Dihydropteroate synthase type 2 (*sul2*) (Morgan et al., 2011). The images of 3D structure models were retrieved from The SWISS-MODEL Repository database (Bienert et al., 2017).

Transformation

Transformation efficiency calculated with positive control, *E. coli* DH5α strain transformed with plasmid pUC19 without inserts, was 4.8 x 10^6 cfu/µg. After electroporation transformant colonies grew on LB agar plates with ampicillin, which indicates that plasmid pUC19 with ampicillin resistance gene was transformed.

The analysis of PCR products with agarose gel electrophoresis revealed right-sized products for each transformant (Fig. 4). However, colony PCR was first attempted with Phusion polymerase for transformants presumably carrying plasmids pUC19, pUC19-*dfrB2_a*, pUC19-*dfrB2_c*, pUC19-*ermB_a* and pUC19-*ermB_b_91N* but with this polymerase it yielded so few products (Fig. 4. A., B., C.), its functionality was questioned. Colony PCR with Phusion polymerase worked sufficiently only for the smallest inserts pUC19, pUC19-*dfrB2_a* and pUC19-*dfrB2_c.* For the rest of the transformants, DreamTaq polymerase was used with good results. This non-proofreading polymerase with lower fidelity gave more reliable results for larger inserts (Fig. 4. D-I).

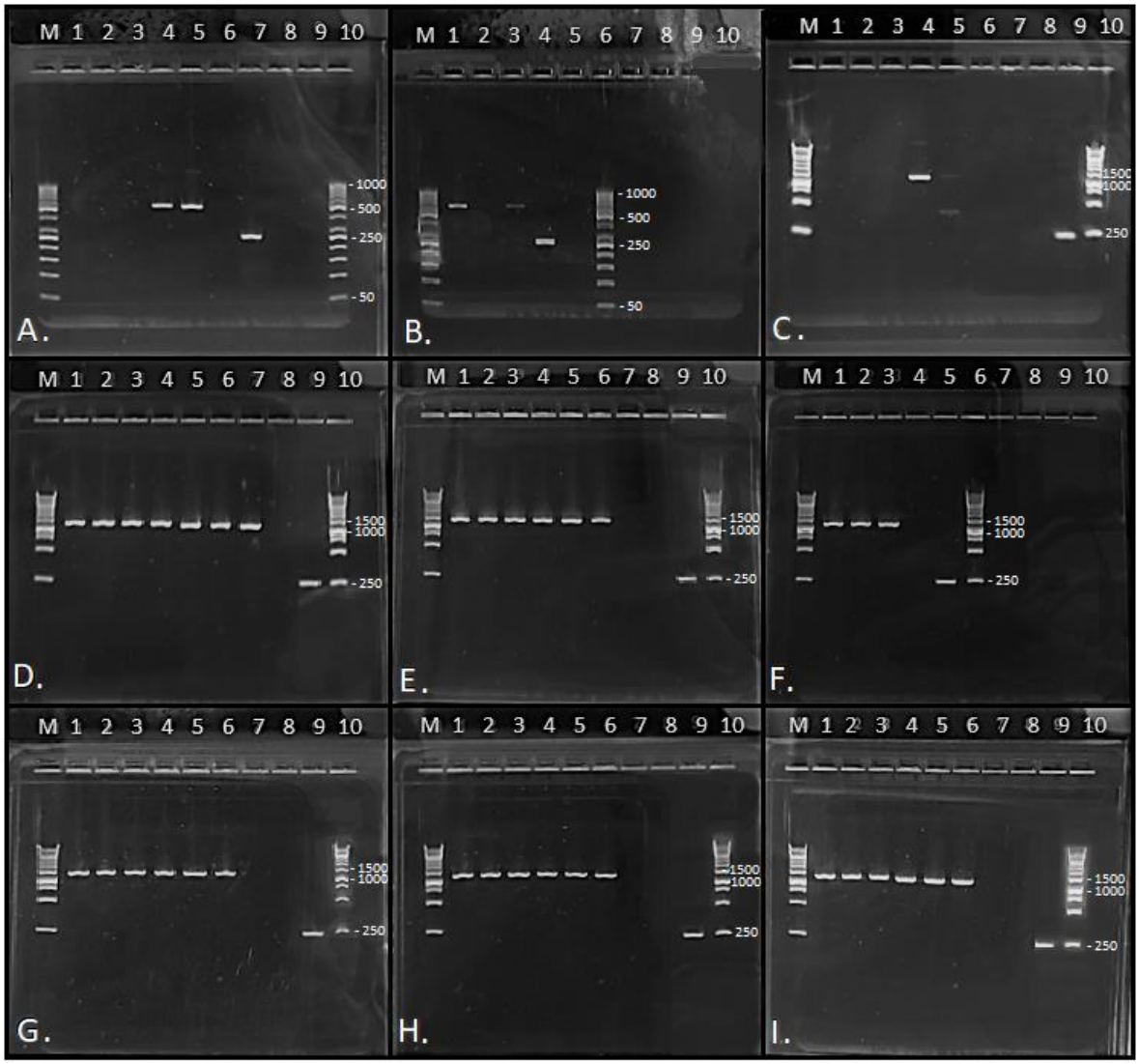

Figure 1. **Agarose gel electrophoresis of colony PCR products.** Molecular marker on gels A and B: GeneRuler 50 bp DNA Ladder. Molecular marker on gels C-I: GeneRuler 1 kb DNA Ladder **A)** 1-3: pUC19-*dfrB2_a* (no product), 4-5: pUC19-*dfrB2_c* (right size product), 6: pUC19-*dfrB2_c* (no product), 7: positive control pUC19 (right size product), 8: negative control. **B)** 1,3: pUC19-*dfrB2_a* (right size product), 2: pUC19-*dfrB2_a* (no product), 4: positive control pUC19 (right size product), 5: negative control. **C)** 1-3: pUC19-*ermB_a* (no product), 4: pUC19-*ermB_b_91N* (right size product), 5-6: pUC19-*ermB_b_91N* (no product), 8: negative control, 9: positive control pUC19 (right size product). **D)** 1-3: pUC19-*ermB_a* (right size product), 4: pUC19-*ermB _b_91N* (right size product), 5-7: pUC19-*ermB_c* (right size product), 8: negative control, 9: positive control pUC19 (right size product). **E)** 1-3: pUC19-*ermC_a* (right size product), 4-6: pUC19-e*rmC_b_72L* (right size product), 8: negative control, 9: positive control pUC19 (right size product). **F)** 1-3: pUC19-*ermC_c* (right size product), 4: negative control, 5: positive control pUC19 (246 bp). **G)** 1-3: pUC19-*sul1_a* (right size product), 4-6: pUC19-*sul1_b_72L* (right size product), 8: negative control, 9: positive control pUC19 (right size product). **H)** 1-3: pUC19-*sul1_c* (right size product), 4-6: pUC19-*sul2_a* (right size product), 8: negative control, 9: positive control pUC19 (right size product). **I)** 1-3: pUC19-*sul2_b_75L* (right size product), 4-6: pUC19-*sul2_c* (right size product), 8: negative control, 9: positive control pUC19 (right size product).

Sequencing data analysis

The results of sequencing data analysis of both pUC19-*ermB_b_91N* extracted from a transformant and the original plasmid delivered by Genscript confirmed the sequences to be as expected. The sequencing products matched with the sequence where *ermB_b_91N* gene is located in pUC19 multiple cloning site. It can be concluded that the results of colony PCR are reliable even though the PCR product sizes seemed slightly bigger on agarose gel analysis.

Antibiotic susceptibility of the strains

Strains DH5α pUC19-*ermB_a* and DH5α pUC19-*ermB_b_91N* displayed equal erythromycin resistance to the wild type strain DH5α pUC19-*ermB_c* (Table 7), indicating that *ermB_a* and *ermB_b_91N* genes conferred resistance to erythromycin similarly as the wild type gene. Strains DH5α pUC19-*ermC_a* and DH5α pUC19-*ermC_b_72L* also displayed erythromycin resistance similar to the wild type strain DH5α pUC19-*ermC_c.*

Strain DH5α pUC19-*sul1_a* gene displayed equal sulfamethoxazole resistance to the wild type strain DH5α pUC19-*sul1_c,* but DH5α pUC19-*sul1_b_72L* strain did not (Table 7). DH5α pUC19-*sul1_b_72L* strain was as sensitive to sulfamethoxazole as the negative control strain. Similarly, DH5α pUC19-*sul2_a* strain displayed equal sulfamethoxazole resistance to the wild type strain DH5α pUC19-*sul2_c* while DH5α pUC19-*sul2_b_75L* strain did not, and the resistance conferred by *sul2_b_75L* was equal to negative control.

The results of the two antibiotic susceptibility tests were highly similar between replicates for strains carrying *ermB*, *ermC*, *sul1* and *sul2* genes and therefore testing was not repeated thrice. The exact MIC values for most of the strains carrying erythromycin and sulfamethoxazole resistance genes could not be determined as the MIC values were observed to be the equal or higher as the highest value on the MIC test strip scale.

Table 7. **Minimal inhibitory concentration (MIC) values of erythromycin and sulfamethoxazole.** MIC values from two tests for strains carrying *ermB*, *ermC*, *sul1* and *sul2* genes, and mean MIC values of the two test results.

Strain	**MIC (µg/ml) of erythromycin**		**MIC (µg/ml) of sulfamethoxazole**		**Mean MIC value (µg/ml)**
	1. test	2. test	1. test	2. test	
DH5α pUC19-*ermB_a*	≥256	≥256			≥256
DH5α pUC19-*ermB_b_91N*	≥256	≥256			≥256
DH5α pUC19-*ermB_c* (WT)	≥256	≥256			≥256
DH5α pUC19-*ermC_a*	≥256	≥256			≥256
DH5α pUC19-*ermC_b_72L*	≥256	≥256			≥256
DH5α pUC19-*ermC_c* (WT)	≥256	≥256			≥256
DH5α pUC19-*sul1_a*			≥1024	1024	≥1024
DH5α pUC19-*sul1_b_72L*			1.5	1.5	1.5
DH5α pUC19-*sul1_c* (WT)			≥1024	≥1024	≥1024
DH5α pUC19-*sul2_a*			≥1024	≥1024	≥1024
DH5α pUC19-*sul2_b_75L*			3	2	2.5
DH5α pUC19-*sul2_c* (WT)			≥1024	≥1024	≥1024
DH5α pUC19 (negative control)	32	32	1.5	1.5	32 (erythromycin) 1.5 (sulfamethoxazole)

DH5α pUC19-*dfrB2_a* strain was more sensitive to trimethoprim than the wild type strain DH5α pUC19-*dfrB2_c* (Table 8). The resistance conferred by *dfrB2_a* was almost equal to negative control strain. The MIC values of trimethoprim for strains DH5α pUC19-*dfrB2_a* and DH5α pUC19-*dfrB2_c* observed in the first two tests differed between replicates and for that reason the test was repeated once more. The MIC values for strain DH5α pUC19-*dfrB2_c* differed the most as can be observed from standard deviation (Table 8). For strain DH5α pUC19-*dfrB2_a* and for the negative control strain the MIC values were more consistent. The MIC values of trimethoprim for strain DH5α pUC19-*dfrB2_c* were the most difficult to interpret in this study due to indistinct inhibition zones (Fig. 5).

Table 8. **Minimal inhibitory concentration (MIC) values of trimethoprim.** MIC values from two tests for strains carrying *dfrB2* gene, and mean MIC values and standard deviations of the three test results.

Strain	**MIC (µg/ml) of trimethoprim**			**Mean MIC value (µg/ml)**	**Standard deviation (µg/ml)**
	1. test	2. test	3. test		
DH5α pUC19-*dfrB2_a*	0.016	0.047	0.032	0.030	0.013
DH5α pUC19-*dfrB2_c* (WT)	0.064	0.190	0.125	0.142	0.051
DH5α pUC19 (negative control)	0.016	0.023	0.032	0.024	0.007

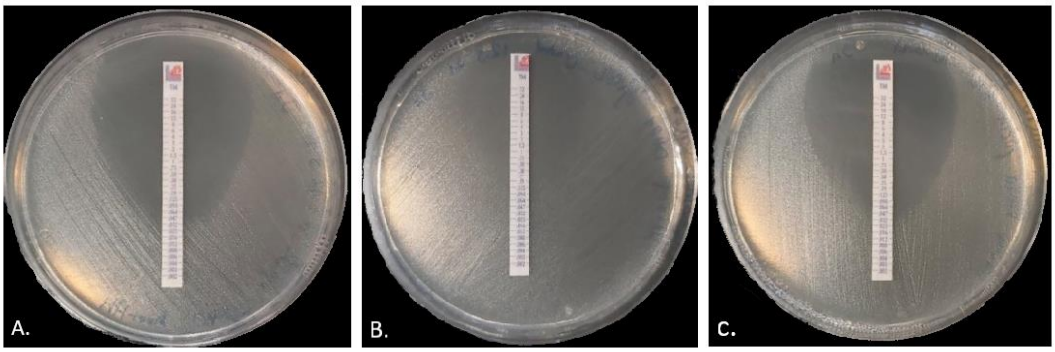

Figure 2. **Inhibition zones of strains DH5α pUC19-*dfrB2_a* (A), DH5α pUC19-*dfrB2_c* (wild type strain) (B) and DH5α pUC19 (negative control strain) (C).** The inhibition zone of DH5α pUC19-*dfrB2_c* was more indistinct than of the other two strains carrying *dfrB2* gene.

The mean MIC values of strains carrying modified ARGs were compared to mean MIC values of strains carrying a wild type gene (Table 7 and 8). No statistical analysis was conducted for antibiotic susceptibility testing results, since the goal was to find out, whether the modified genes function well enough compared to the wild type or not.

In conclusion, six of the ARGs with internal modifications (*ermB_a, ermB_b_91N, ermC_a, ermC_b_72L, sul1_a* and *sul2_a*) conferred antibiotic resistance similarly to the wild type gene while three (*dfrB2_a*, *sul1_b_72L* and *sul2_b_75L*) did not. Two out of four (*ermB_b_91N* and *ermC_b_72L*) ARGs with insertions in predicted permissive stretches conferred antibiotic resistance equally to the wild type gene. S*ul1_b_72L* and *sul2_b_75L* did not confer antibiotic resistance. Fig. 6 summarizes the results of antibiotic susceptibility testing by illustrating the antibiotic resistance levels conferred by the modified ARGs compared to the wild type. The accurate values are presented in supplementary Table S1.

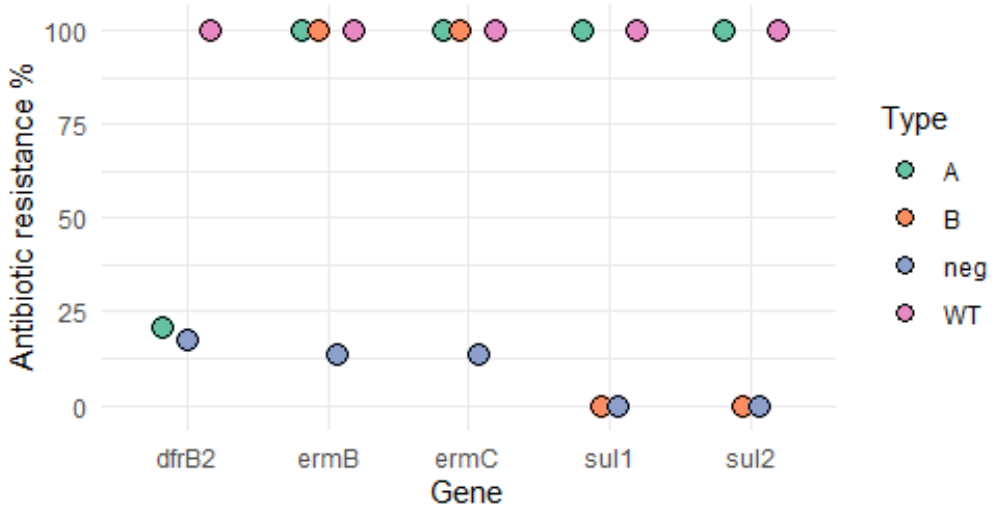

Figure 6. **The antibiotic resistance levels conferred by modified antibiotic resistance genes compared to wild type gene (WT).** The wild type gene is selected as the 100% level and modified antibiotic resistance genes and negative control are compared to it. The explanations of the types of genes (A and B) are in Fig. 1.

DISCUSSION

The aim of this study was to find out whether synthetic primer binding sites and barcode sequence can be introduced inside various ARGs without greatly disturbing protein function. The goal of antibiotic susceptibility testing was to determine whether the modified ARGs encode for functional proteins conferring antibiotic resistance, and whether the antibiotic resistance of the strains carrying the modified ARGs is similar enough to a strain carrying a wild type ARG. The functionality of the protein was to be maintained despite modifications to allow the ARGs to be selected by a corresponding antibiotic and thus to have applicable ARGs for further studies where the HGT of the genes will be monitored by tracing the ARGs with epicPCR. It can be concluded that six ARGs modified in this study are adequate for further studies since they confer sufficient antibiotic resistance providing advantage to the host strain and are therefore selectable with the corresponding antibiotic. Moreover, the epicPCR primer binding sites could be added inside these ARGs without impairing protein function, which allows the monitoring of ARG transfer in experimental conditions. The results of this study indicate that the *ermB_a, ermB_b_91N, ermC_a, ermC_b_72L* as well as *sul1_a* and *sul2_a* genes encode for functional proteins conferring antibiotic

resistance to erythromycin or sulfamethoxazole respectively. Due to the limited MIC test strip scale, it could not be determined whether the ARGs confer resistance equal to wild type gene, yet this was not the aim of this study. The modified ARGs are intended for an experimental evolution experiment, where they need to bring a significant advantage for their host. Therefore, the modified ARGs do not need to be equal to wild type in terms of conferring antibiotic resistance.

Clinical and Laboratory Standards Institute (CLSI) has published MIC breakpoints of trimethoprim and sulfonamides for Enterobacterales, which include *E. coli* (Clinical and Laboratory Standards Institute, 2021). However, there are no MIC breakpoint values of erythromycin for Enterobacterales, since erythromycin is not commonly used to treat infections caused by them. Strains DH5α pUC19-*sul1_a,* DH5α pUC19-*sul1_c,* DH5α pUC19-*sul2_a* and DH5α pUC19-*sul2_c* can be categorized as resistant to sulfamethoxazole as the determined MIC values of the strains exceed the MIC breakpoint defined by CLSI (more than 512 µg/ml), whereas strains with *sul1_ b_72L* and *sul2_b_75L* genes can be categorized as sensitive (determined MIC values are less than 256 µg/ml). The fact that the strains carrying *sul1_a* and *sul2_a* genes are resistant to sulfamethoxazole according to MIC breakpoints defined by CLSI supports the conclusion that the function of proteins encoded by genes *sul1_a* and *sul2_a* was maintained despite modifications and that the epicPCR primer binding site can be added inside start and stop codons for these genes. According to MIC breakpoint values defined by CLSI, the strains carrying modified and wild type *dfrB2* would not be resistant to trimethoprim at all, since the MIC values determined in this study for these strains are considerably less than the MIC breakpoint by CLSI (more than 16 µg/ml for the category “resistant”). Both the modified strain DH5α pUC19-*dfrB2_a* as well as the wild type strain DH5α pUC19-*dfrB2_c* can be categorized as sensitive (determined MIC values are less than 8 µg/ml). Although the wild type DH5α pUC19-*dfrB2_c* strain was found not to be resistant to trimethoprim according to the MIC breakpoints by CLSI, clearly higher MIC values were determined for wild type DH5α pUC19-*dfrB2_c* strain than for the negative control strain or the modified DH5α pUC19-*dfrB2_a* strain. This allowed comparison of resistance levels, however with reservations since the results also differed between replicates. Since *E. coli* is a known host of *dfrB2* gene and thus the gene should function properly, the reason why the wild type *dfrB2_c* conferred resistance poorly remains unclear. A point mutation in the gene sequence could explain the poor functionality and plasmid sequencing would provide more information on this. Noteworthy is also the addition of ampicillin to the plates in antibiotic

susceptibility testing to avoid a loss of plasmid. The additional antibiotic on a plate also containing a MIC test strip impregnated with another antibiotic may have caused an extra burden for bacteria resulting in impaired growth of strains with both *dfrB2* genes.

In this study, all the other proteins with C- and N-terminal modifications (type A genes) maintained their function, but DH5α pUC19-*dfrB2_a* strain was observed to not confer trimethoprim resistance, indicating a loss of function of the protein encoded by *dfrB2_a*. The 3D structure of protein encoded by *dfrB2* gene was not evaluated, as the modifications were done at the C- and N-terminal ends of the protein and modifying the ends of the protein is safe for many proteins in terms of maintaining protein function. However, it is possible that a terminal region is also functionally relevant or that the insertion at an end interfered with protein folding.

Based on the results of this study, sequence-based prediction of permissive sites seems to be a useful method, but it does not guarantee the functionality of the engineered protein. As observed in the antibiotic susceptibility testing, two out of four proteins with insertions in predicted permissive stretches maintained their function, while the other two lost their ability to confer antibiotic resistance. Possible reasons why the engineered proteins with insertions in predicted permissive stretches did not function in this study include that the region of insertion was functionally relevant or that the insertion interfered with protein folding thus impairing the function. Loop regions where the insertions were aimed are generally thought to be flexible, but they can also have a functional role (Papaleo et al., 2016). For example, surface-exposed loops can be functionally relevant due to their possibility to interact with biomolecules. Therefore, placing insertions on a loop structure does not guarantee that a functional region will not be interfered. Another possible reason for the loss of protein function is the inaccuracy in identifying the permissive site. The position of the permissive stretch identified by the gap in MSA is not exact as the used algorithm can affect the precise location of the gap (Golubchik et al., 2007). Exploring multiple insertion sites in a protein (Burg et al., 2016; Oesterle et al., 2017; Schlehuber & Rose, 2004) might give a better chance of resulting in a functional protein, although a good result can be achieved with one try (Sturgill et al., 2008). One predicted permissive site for insertion was tested for proteins of interest in this study. However, it would have been justified to try more than one insertion site, as well as to vary the insertion site within the identified stretch. Especially in the case of the protein encoded by *sul1_ b_72L*, placing the insertion further away from the secondary

structure and towards the mid of the loop structure might be worth exploring. Since neither of proteins encoded by genes *sul1_ b_72L* and *sul2_b_75L* was functional, it is possible that the insertions done in this study for these proteins were too close to the secondary structure.

To minimize the risk of insertions in a location that would impair the function of the modified protein, it might be useful to utilize a protein secondary structure prediction tool such as JPred4 in the future. JPred4 has a secondary structure prediction accuracy of around 80% (Drozdetskiy et al., 2015). Still, because a 3D model of the protein structure is more useful than information on secondary structure only, using a prediction tool might not have changed the course of this study. It has also been suggested that sequence-based prediction for identification of permissive sites does not need pre-existing data at all on functionally important regions or on protein structures (Burg et al., 2016). The 3D structure models of each protein were nevertheless evaluated in this study to place the insertions in loop regions.

The MIC values obtained by antimicrobial gradient method in this study cannot be considered as exact MIC values for each strain. Yet for the purposes of this study, the antimicrobial gradient method is accurate enough as the observed MIC values are comparable with each other. The antimicrobial gradient method was chosen for this study based on its fast and easy performance compared to other antibiotic susceptibility testing methods such as the broth dilution method. The antimicrobial gradient method has limitations compared to other methods but performs generally well (Jorgensen & Ferraro, 2009) and it is accurate enough for clinical use (Baker et al., 1991). A commercial version of antimicrobial gradient method called Etest® (BioMérieux) is shown in several studies to be an accurate alternative for broth dilution or agar dilution methods (Baker et al., 1991; Di Bonaventura et al., 2002; Ming Bo Huang et al., 1992). Studies comparing Etest® and Liofilchem® MIC test strips used in this study have shown that Liofilchem® MIC test strips also provide acceptable results, but they seem to be slightly less accurate than those obtained by Etest® (Humphries et al., 2018; Jönsson et al., 2018). The MIC test strips used in this study were from the same batch. Should there be deficiencies in the strips of this batch, it will show in all the results. The use of strips from the same batch therefore exposes the results to random deficiencies in manufacturing, but on the other hand the comparability of the results is maintained. Also, the antimicrobial gradient method is based on visual interpretation of results which may lead to dependence of the results on the performer of the test. Moreover, the true MIC values for strains carrying erythromycin and

sulfamethoxazole resistance genes could not be determined as the MIC values were observed to be the equal or higher as the highest value on the MIC test strip scale.

In this study, *E. coli* DH5α strain used as a host and as a negative control was observed to have an innate resistance against erythromycin. This is consistent with previous studies analyzing clinical isolates where high erythromycin resistance levels in different *E. coli* strains have been observed (Kazemnia et al., 2014; Kibret & Abera 2011). Despite the natural erythromycin resistance of the host strain, the differences in MIC values between the negative control and the tested strains were clearly observable. In the future, further studies with *ermB* and *ermC* genes should be done by using for example erythromycin sensitive strain of *Staphylococcus aureus* as a host. Most methicillin sensitive strains of *S. aureus* are erythromycin sensitive (Cheng et al., 2016; Kareiviene et al., 2006) but *S. aureus* is known to express *erm*-mediated erythromycin resistance acquired via plasmids and other mobile genetic elements (Haaber et al., 2017). The ability to express *erm*-mediated erythromycin resistance acquired via plasmids suits the purposes of further studies, but when choosing a suitable strain, it should be checked that the strain is not already resistant. Use of *S. aureus* would also require a different plasmid since pUC19 does not replicate in it.

The ARGs in which the modifications were done at the ends of the gene (type A), result in an epicPCR product longer than the desired 300 bp. In the future the applicability of these genes to epicPCR needs to be confirmed. Another future perspective is determining the effect of different barcode sequences on protein function. In this study, the same barcode sequence after the epicPCR forward primer binding site was used in all the genes to avoid different result in antibiotic susceptibility testing caused by different barcodes sequences.

In conclusion, six ARGs modified in this study encoded for functional proteins that conferred antibiotic resistance while three modified ARGs did not. According to the results, the epicPCR primer binding sites could be added inside the genes without impairing protein function, making the ARGs selectable by corresponding antibiotic and amplifiable by epicPCR. The functional, antibiotic resistance conferring genes designed in this study can be used in further studies utilizing epicPCR, provided that the length of the gene sequence is not too long for epicPCR. In addition, the results of this study provide information on the usability of sequence-based prediction in finding permissive stretches for internal insertions of proteins.

ACKNOWLEDGEMENTS

I would like to thank both of my supervisors Veera Partanen and Marko Virta for their excellent supervision, scientific guidance, and support. I would also like to express my gratitude to the whole Molecular Environmental Biosciences group in University of Helsinki for the supporting working environment and for giving me valuable advice and feedback during this project.

REFERENCES

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J. & Handelsman, J. 2010, "Call of the wild: Antibiotic resistance genes in natural environments", *Nature Reviews Microbiology*, vol. 8, no. 4, pp. 251-259.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990, "Basic local alignment search tool", *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403-410.

Baker, C.N., Stocker, S.A., Culver, D.H. & Thornsberry, C. 1991, "Comparison of the E test to agar dilution, broth microdilution, and agar diffusion susceptibility testing techniques by using a special challenge set of bacteria", *Journal of Clinical Microbiology*, vol. 29, no. 3, pp. 533-538.

Bateman, A., Martin, M., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., … Zhang, J. & UniProt Consortium 2021, "UniProt: The universal protein knowledgebase in 2021", *Nucleic Acids Research,* vol. 49, no. D1, pp. D480-D489.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. 2000, "The Protein Data Bank", *Nucleic Acids Research*, vol. 28, no. 1, pp. 235-242.

Bhujbalrao, R. & Anand, R. 2019, "Deciphering determinants in ribosomal methyltransferases that confer antimicrobial resistance", *Journal of the American Chemical Society*, vol. 141, no. 4, pp. 1425-1429.

Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L. & Schwede, T. 2017, "The SWISS-MODEL Repository-new features and functionality", *Nucleic Acids Research*, vol. 45, no. D1, pp. D313-D319.

Billerbeck, S., Calles, B., Müller, C.L., De Lorenzo, V. & Panke, S. 2013, "Towards functional orthogonalisation of protein complexes: Individualisation of GroEL monomers leads to distinct quasihomogeneous single rings", *ChemBioChem*, vol. 14, no. 17, pp. 2310-2321.

Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J. & Madden, T.L. 2012, "Domain enhanced lookup time accelerated BLAST", *Biology Direct*, vol. 7.

Boutet E., Lieberherr D., Tognolli M., Schneider M. & Bairoch A. 2007. UniProtKB/Swiss-Prot. *Methods in Molecular Biology*, vol. 406, pp. 89-112.

Burg, L., Zhang, K., Bonawitz, T., Grajevskaja, V., Bellipanni, G., Waring, R. & Balciunas, D. 2016, “Internal epitope tagging informed by relative lack of sequence conservation”, *Scientific Reports*, vol. 6.

Cairns, J., Ruokolainen, L., Hultman, J., Tamminen, M., Virta, M. & Hiltunen, T. 2018, "Ecology determines how low antibiotic concentration impacts community composition and horizontal transfer of resistance genes", *Communications Biology*, vol. 1, no. 1.

Chang, M.S.S. & Benner, S.A. 2004, "Empirical analysis of protein insertions and deletions determining parameters for the correct placement of gaps in protein sequence alignments", *Journal of Molecular Biology*, vol. 341, no. 2, pp. 617-631.

Cheng, M.P., René, P., Cheng, A.P. & Lee, T.C. 2016, "Back to the Future: Penicillin-Susceptible *Staphylococcus aureus*", *American Journal of Medicine*, vol. 129, no. 12, pp. 1331-1333.

Clinical and Laboratory Standards Institute 2021. Performance Standards for Antimicrobial Susceptibility Testing. 31st Edition. CLSI guideline M100. Wayne, PA: Clinical and Laboratory Standards Institute, USA.

Di Bonaventura, G., D'Antonio, D., Catamo, G., Ballone, E. & Piccolomini, R. 2002, "Comparison of Etest, agar dilution, broth microdilution and disk diffusion methods for testing in vitro activity of levofloxacin against *Staphylococcus* spp. isolated from neutropenic cancer patients", *International Journal of Antimicrobial Agents*, vol. 19, no. 2, pp. 147-154.

Drozdetskiy, A., Cole, C., Procter, J. & Barton, G.J. 2015, "JPred4: A protein secondary structure prediction server", *Nucleic Acids Research*, vol. 43, no. W1, pp. W389-W394.

Golubchik, T., Wise, M.J., Easteal, S. & Jermiin, L.S. 2007, "Mind the gaps: Evidence of bias in estimates of multiple sequence alignments", *Molecular Biology and Evolution*, vol. 24, no. 11, pp. 2433-2442.

Goodale, A., Michailidis, F., Watts, R., Chok, S.C. & Hayes, F. 2020, "Characterization of permissive and non-permissive peptide insertion sites in chloramphenicol acetyltransferase", *Microbial Pathogenesis*, vol. 149.

Haaber, J., Penadés, J.R. & Ingmer, H. 2017, "Transfer of antibiotic resistance in *Staphylococcus aureus*", *Trends in Microbiology*, vol. 25, no. 11, pp. 893-905.

Hernando-Amado, S., Coque, T.M., Baquero, F. & Martínez, J.L. 2019, "Defining and combating antibiotic resistance from One Health and Global Health perspectives", *Nature Microbiology*, vol. 4, no. 9, pp. 1432-1442.

Hiltunen, T., Virta, M. & Laine, A. 2017, "Antibiotic resistance in the wild: An ecoevolutionary perspective", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1712.

Hughes, V.M. & Datta, N. 1983, "Conjugative plasmids in bacteria of the 'pre-antibiotic' era ", *Nature*, vol. 302, no. 5910, pp. 725-726.

Hultman, J., Tamminen, M., Pärnänen, K., Cairns, J., Karkman, A. & Virta, M. 2018, "Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent", *FEMS Microbiology Ecology*, vol. 94, no. 4.

Humphries, R.M., Hindler, J.A., Magnano, P., Wong-Beringer, A., Tibbetts, R. & Miller, S.A. 2018, "Performance of ceftolozane-tazobactam etest, MIC test strips, and disk diffusion compared to reference broth microdilution for β-Lactam-Resistant *Pseudomonas aeruginosa* isolates", *Journal of Clinical Microbiology*, vol. 56, no. 3.

International.neb.com, 2015. *Monarch® Plasmid DNA Miniprep Kit Protocol (NEB #T1010).* [online] Available at: https://international.neb.com/protocols/2015/11/20/monarch-plasmid-dna-miniprep-kit-protocol-t1010 [Accessed 17 Feb. 2021].

Jorgensen, J.H. & Ferraro, M.J. 2009, "Antimicrobial susceptibility testing: A review of general principles and contemporary practices", *Clinical Infectious Diseases*, vol. 49, no. 11, pp. 1749-1755.

Jönsson, A., Jacobsson, S., Foerster, S., Cole, M.J. & Unemo, M. 2018, "Performance characteristics of newer MIC gradient strip tests compared with the Etest for antimicrobial susceptibility testing of *Neisseria gonorrhoeae*", *APMIS*, vol. 126, no. 10, pp. 822-827.

Kareiviene, V., Pavilonis, A., Sinkute, G., Liegiute, S. & Gailiene, G. 2006, "*Staphylococcus aureus* resistance to antibiotics and spread of phage types.", *Medicina (Kaunas, Lithuania)*, vol. 42, no. 4, pp. 332-339.

Kazemnia, A., Ahmadi, M. & Dilmaghani, M. 2014, "Antibiotic resistance pattern of different *Escherichia coli* phylogenetic groups isolated from human urinary tract infection and avian colibacillosis", *Iranian Biomedical Journal*, vol. 18, no. 4, pp. 219-224.

Kibret, M. & Abera, B. 2011, "Antimicrobial susceptibility patterns of *E. coli* from clinical sources in northeast Ethiopia", *African Health Sciences*, vol. 11, no. SPEC. ISSUE, pp. S40-S45.

MacArthur, M.W. & Thornton, J.M. 1991, "Influence of proline residues on protein conformation", *Journal of Molecular Biology*, vol. 218, no. 2, pp. 397-412.

Matsumura, M., Signor, G. & Matthews, B.W. 1989, "Substantial increase of protein stability by multiple disulphide bonds", *Nature*, vol. 342, no. 6247, pp. 291-293.

Ming Bo Huang, Baker, C.N., Banerjee, S. & Tenover, F.C. 1992, "Accuracy of the E test for determining antimicrobial susceptibilities of staphylococci, enterococci, *Campylobacter jejuni*, and gram-negative bacteria resistant to antimicrobial agents", *Journal of Clinical Microbiology*, vol. 30, no. 12, pp. 3243-3248.

Morgan, R.E., Batot, G.O., Dement, J.M., Rao, V.A., Eadsforth, T.C. & Hunter, W.N. 2011, "Crystal structures of *Burkholderia cenocepacia* dihydropteroate synthase in the apo-form and complexed with the product 7,8-dihydropteroate", *BMC Structural Biology*, vol. 11.

New England Biolabs 2019, "NEBuilder® HiFi DNA Assembly Master Mix/NEBuilder HiFi DNA Assembly Cloning Kit", Instruction Manual, Version 2.0 8/19.

Oesterle, S., Roberts, T.M., Widmer, L.A., Mustafa, H., Panke, S. & Billerbeck, S. 2017, "Sequence-based prediction of permissive stretches for internal protein tagging and knockdown", *BMC Biology*, vol. 15

Papaleo, E., Saladino, G., Lambrughi, M., Lindorff-Larsen, K., Gervasio, F.L. & Nussinov, R. 2016, "The role of protein loops and linkers in conformational dynamics and allostery", *Chemical Reviews,* vol. 116, no. 11, pp. 6391-6423.

Rasmussen, J.L., Odelson, D.A. & Macrina, F.L. 1986, "Complete nucleotide sequence and transcription of *ermF*, a macrolide-lincosamide-streptogramin B resistance determinant from *Bacteroides fragilis*", *Journal of Bacteriology*, vol. 168, no. 2, pp. 523-533.

Rizzo, L., Manaia, C., Merlin, C., Schwartz, T., Dagot, C., Ploy, M.C., Michael, I. & Fatta-Kassinos, D. 2013, "Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review", *Science of the Total Environment*, vol. 447, pp. 345-360.

Schlehuber, L.D. & Rose, J.K. 2004, "Prediction and identification of a permissive epitope insertion site in the vesicular stomatitis virus glycoprotein", *Journal of Virology*, vol. 78, no. 10, pp. 5079-5087.

Schluckebier, G., Zhong, P., Stewart, K.D., Kavanaugh, T.J. & Abad-Zapatero, C. 1999, "The 2.2 Å structure of the rRNA methyltransferase *ErmC*' and its complexes with cofactor and cofactor analogs: Implications for the reaction mechanism", *Journal of Molecular Biology*, vol. 289, no. 2, pp. 277-291.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. & Higgins, D.G. 2011, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega", *Molecular Systems Biology*, vol. 7.

Spencer, S.J., Tamminen, M.V., Preheim, S.P., Guo, M.T., Briggs, A.W., Brito, I.L., A Weitz, D., Pitkänen, L.K., Vigneault, F., Virta, M.P. & Alm, E.J. 2016, "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers", *ISME Journal*, vol. 10, no. 2, pp. 427-436.

Sturgill, T.W., Cohen, A., Diefenbacher, M., Trautwein, M., Martin, D.E. & Hall, M.N. 2008, "TOR1 and TOR2 have distinct locations in live cells", *Eukaryotic Cell*, vol. 7, no. 10, pp. 1819-1830.

Thiele-Bruhn, S. 2003, "Pharmaceutical antibiotic compounds in soils - A review", *Journal of Plant Nutrition and Soil Science*, vol. 166, no. 2, pp. 145-167.

World Health Organization, 2020. *Antibiotic resistance*. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance [Accessed 21 September 2021].

Yanisch-Perron, C., Vieira, J. & Messing, J. 1985, "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors", *Gene,* vol. 33, no. 1, pp. 103-119.

Yun, M.-K., Wu, Y., Li, Z., Zhao, Y., Waddell, M.B., Ferreira, A.M., Lee, R.E., Bashford, D. & White, S.W. 2012, "Catalysis and sulfa drug resistance in dihydropteroate synthase", *Science*, vol. 335, no. 6072, pp. 1110-1114.

SUPPORTING DATA

Table S1. **The accurate values of antibiotic resistance levels conferred by modified antibiotic resistance genes compared to wild type gene (WT).** The wild type gene is selected as the 100% level and modified antibiotic resistance genes and negative control are compared to it.

Strain	**Antibiotic resistance (%) compared to wild type strain**
DH5α pUC19 + *dfrB2_a*	20.74 %
DH5α pUC19 + *dfrB2_c* (WT)	100.00 %
DH5α pUC19 (neg)	16.64 %
DH5α pUC19-*ermB_a*	100.00 %
DH5α pUC19-*ermB_b_91N*	100.00 %
DH5α pUC19-*ermB_c* (WT)	100.00 %
DH5α pUC19 (neg)	12.50 %
DH5α pUC19-*ermC_a*	100.00 %
DH5α pUC19-*ermC_b_72L*	100.00 %
DH5α pUC19-*ermC_c* (WT)	100.00 %
DH5α pUC19 (neg)	12.50 %
DH5α pUC19-*sul1_a*	100.00 %
DH5α pUC19-*sul1_b_72L*	0.15 %
DH5α pUC19-*sul1_c* (WT)	100.00 %
DH5α pUC19 (neg)	0.15 %
DH5α pUC19-*sul2_a*	100.00 %
DH5α pUC19-*sul2_b_75L*	0.24 %
DH5α pUC19-*sul2_c* (WT)	100.00 %
DH5α pUC19 (neg)	0.15 %